ECON 5033 Econometrics I – Lecture 4

Multiple Linear Regression Model – Part II

Eric S. Lin

Department of Economics, National Tsing Hua University

October 15, 2024

An Additional Assumption

- Recall the Assumptions in multiple linear regression model:
  1. $\mathrm{E}[Y|X] = X\beta$
  2. $\mathrm{E}[\varepsilon|X] = 0$
  3. $\mathrm{E}[\varepsilon\varepsilon'|X] = \sigma^2 I_n$
  4. $\mathrm{rank}(X) = k$
  5. $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$
- To explore the exact sampling distribution of OLS estimators, we need the Assumption 5.
- Given Assumptions A1-A5, we have the exact sampling distribution of $\hat{\beta}$

$$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 \left(X'X\right)^{-1}).$$

Sampling Distribution of Error Variance

- In the previous lecture we've proposed an (unbiased) estimator of σ^2:

$$\hat{\sigma}^2 = \frac{e'e}{n-k}.$$

- The sampling distribution of $\hat{\sigma}^2$ is given by the following theorem.

Theorem

$(n-k)\,\hat{\sigma}^2/\sigma^2 \sim \mathcal{X}^2\,(n-k)\,.$

(Proof) Sketch:

$(n-k)\,\hat{\sigma}^2/\sigma^2 = \varepsilon' M_X \varepsilon/\sigma^2$. Note that $M_X = C'\Lambda C$ and $C'C = I$, where Λ is a diagonal matrix with $(n-k)$ eigenvalues of 1's and k eigenvalues of 0's. Let $z = \varepsilon/\sigma \sim \mathcal{N}\,(0, I_n)\,.$ Then, $\varepsilon' M_X \varepsilon/\sigma^2 = \sum_{j=1}^{n-k} z_j^2$. □

Sampling Distribution of Error Variance

- We can also show that $\hat{\sigma}^2$ (a quadratic form) is independent of $\hat{\beta}$ (a linear function) but the following lemma is needed.
- The *t*-distribution will be used in many forms of hypothesis tests, which can be expressed as the ratio of a linear to quadratic form in a normal vector.
  - Before that, we have a linear to linear independence result.

Lemma

If $X \sim \mathcal{N}(0, I_n)$ and if $AB' = 0$, then AX and BX are independent.

Proof.

Since $\text{cov}(AX, BX) = AB' = 0$ □

Sampling Distribution of Error Variance

- Now we present the relationship between a linear function and a quadratic form.

Lemma

If the random vector (of dimension n) $X \sim \mathcal{N}(0, I_n)$, then AX and a quadratic form $X'BX$ (symmetric and idempotent B) are independent if $AB' = 0$.

Proof.

Write $X'BX = (XB)'BX$. It follows that $\mathrm{cov}(AX, BX) = 0$ since $AB' = 0$. □

Sampling Distribution of Error Variance

- Can extend to two quadratic forms for constructing a F-distribution

Lemma

If the random vector (of dimension n) $X \sim \mathcal{N}(0, I_n)$, then $X'AX$ and $X'BX$ (symmetric and idempotent A and B) are independent if $AB' = 0$.

Sampling Distribution of Error Variance

Theorem

$\hat{\sigma}^2$ *and* $\hat{\beta}$ *are independent.*

Proof:

The result follows directly by the above lemma. □

Single Coefficient Test

- In this case, it is pretty much the same as we learned before in simple linear regression model.
- Consider the following hypothesis:

$$
\begin{aligned}
H_o &: \beta_j = \beta_{jo} \in \mathcal{R} \\
H_1 &: \beta_j \neq \beta_{jo}
\end{aligned}
$$

- The form of the *t*-test is:

$$
\frac{\hat{\beta}_j - \beta_{jo}}{\sqrt{\hat{\sigma}^2 \left\{ (X'X)^{-1} \right\}_{jj}}} \sim t\,(n-k)\,.
$$

Simple Linear Test

Example

In Greene's textbook, the following linear regression model is established to describe the price of painting in an auction:

$$\ln P_i = \beta_1 + \beta_2 \ln S_i + \beta_3 AR_i + \varepsilon_i,$$

where P_i is the price of a painting, S_i is its size, AR_i denotes "aspect ratio". We are not quite sure that this model is correct, because it is questionable whether the size of a painting affects its price. For instance, Mona Lisa by Leonardo da Vinci is very small-sized. We can test the Hypothesis by:

$$H_o : \beta_2 = 0.$$

Simple Linear Test

- Sometimes we need to test the single linear hypothesis as follows.

$$H_o : c'\beta = r \in \mathcal{R}$$
$$H_1 : c'\beta \neq r$$

- Under H_o, it is straightforward to have

$$(c'\hat{\beta} - r)|X \sim \mathcal{N}(0, \sigma^2 c' \left(X'X\right)^{-1} c)$$
$$\frac{c'\hat{\beta} - r}{\sqrt{\sigma^2 c' \left(X'X\right)^{-1} c}}|X \sim \mathcal{N}\left(0, 1\right).$$

- A nuisance parameter here, i.e., under H_o, it is not a pivotal statistic.

Simple Linear Test

- Idea is to replace σ^2 by $\hat{\sigma}^2$ to form a t-statistic:

$$\frac{c'\hat{\beta} - r}{\sqrt{\hat{\sigma}^2 c' \left(X'X\right)^{-1} c}} = \frac{\frac{c'\hat{\beta} - r}{\sqrt{\sigma^2 c'(X'X)^{-1} c}}}{\sqrt{\left(n-k\right)\hat{\sigma}^2 \sigma^{-2} / \left(n-k\right)}} = \frac{\text{term I}}{\text{term II}} \tag{1}$$

- What about term I?
- What about term II?
- term I/ termII is distributed as a t distribution by ...
- (1) is pivotal now.
- If $|t| > t_{n-k,\alpha/2}$, one could reject the null.

Simple Linear Test

Example

Consider the Cobb-Douglas production model, $Y_i = AK_i^{\beta_2} L_i^{\beta_3} e^{\varepsilon_i}$. Taking logs yields our linear regression model:

$$y_i = \beta_1 + \beta_2 k_i + \beta_3 l_i + \varepsilon_i.$$

Hypothesis of constant return to scale (CRS) is given by:

$$H_o : \beta_2 + \beta_3 = 1 \ (c'\beta = 1) \text{ vs. } H_1 : \beta_2 + \beta_3 \neq 1,$$

where $c' = (0, 1, 1)$, $\beta = (\beta_1, \beta_2, \beta_3)'$ and $r = 1$. Test statistic is:

$$t = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 1}{s.e.(\hat{\beta}_2 + \hat{\beta}_3 - 1)} = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 1}{s.e.(\hat{\beta}_2 + \hat{\beta}_3)} \sim t(n-3).$$

Simple Linear Test

Example (Continued)

Rewrite our linear regression model:

$$
\begin{aligned}
y_i &= \beta_1 + \beta_2 k_i + \beta_3 l_i + \varepsilon_i \\
&= \beta_1 + \beta_2 k_i + \beta_3 k_i - \beta_3 k_i + \beta_3 l_i + \varepsilon_i \\
&= \beta_1 + (\beta_2 + \beta_3) k_i + \beta_3 (l_i - k_i) + \varepsilon_i \\
&= \beta_1 + \beta_2^* k_i + \beta_3 l_i^* + \varepsilon_i
\end{aligned}
$$

Hypothesis of CRS is given by:

$$H_o : \beta_2^* = 1 \text{ vs. } H_1 : \beta_2^* \neq 1.$$

Sometimes this is called re-parameterization.

Joint Linear Test

- In this case, we assume linear constraints to be linearly independent.
- Let R be a fixed known $J \times k$ matrix (assume $J < k$) of full rank J.
- The hypothesis is:

$$H_o : R\beta = q \in \mathcal{R}^J$$
$$H_1 : R\beta \neq q.$$

- Similarly, under H_o, one could have:

$$(R\hat{\beta} - q)|X \sim \mathcal{N}(0, \sigma^2 R \left(X'X\right)^{-1} R')$$

$$(\sigma^2 R \left(X'X\right)^{-1} R')^{-1/2}(R\hat{\beta} - q)|X \sim \mathcal{N}\left(0, I_J\right).$$

- Recall the t or Z test we learned before. Is it a scalar test statistic?

Joint Linear Test

Example

Assume $k = 6$, our model is:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_6 X_{6i} + \varepsilon_i.$$

Consider the restriction $\beta_2 + \beta_3 = 1, \beta_4 - \beta_6 = 0$ and let

$$R = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix} \text{ and } q = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We can then write:

$$H_o : R\beta = q = (1, 0)' \in \mathcal{R}^2.$$

Joint Linear Test

- It turns out that we could form the following.

$$(R\hat{\beta}-q)'(\sigma^2 R\left(X'X\right)^{-1}R')^{-1}(R\hat{\beta}-q)\sim\mathcal{X}^2\left(J\right).$$

- Problem?
- F type statistic seems useful to eliminate the unknown parameters.

$$\begin{aligned}&\frac{(R\hat{\beta}-q)'(\hat{\sigma}^2 R\left(X'X\right)^{-1}R')^{-1}(R\hat{\beta}-q)}{J}\\ &=\frac{(R\hat{\beta}-q)'(R\left(X'X\right)^{-1}R')^{-1}(R\hat{\beta}-q)/J}{e'e/\left(n-k\right)}\\ &=\frac{(R\hat{\beta}-q)'(\sigma^2 R\left(X'X\right)^{-1}R')^{-1}(R\hat{\beta}-q)/J}{\left(n-k\right)\sigma^{-2}\hat{\sigma}^2/\left(n-k\right)}\sim F\left(J,n-k\right)\end{aligned}$$

Joint Linear Test

Example

To test the overall significance of a regression, one could set up the null below.

$$H_o : \beta_2 = \beta_3 = ... = \beta_k = 0 \in \mathcal{R}^{k-1}$$
$$H_1 : \text{at least one of these } \beta\text{'s } \neq 0.$$

In this case, we have

$$R = \left(0_{(k-1)\times 1}, I_{k-1}\right)_{(k-1)\times k}$$
$$q = 0_{(k-1)\times 1}$$
$$X = (X_1, X_2) = (\iota, X_2)$$

Test of the Overall Significance

Example (Continued)

$$X'X = \begin{bmatrix} n & \iota' X_2 \\ X_2'\iota & X_2'X_2 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} \cdot & \cdot \\ \cdot & (X_2' M_\iota X_2)^{-1} \end{bmatrix}$$

$$(R(X'X)^{-1} R')^{-1} = X_2' M_\iota X_2 = x_2' x_2$$

$$R\hat{\beta} - q = \hat{\boldsymbol{\beta}}_2$$

Based on above information, we have

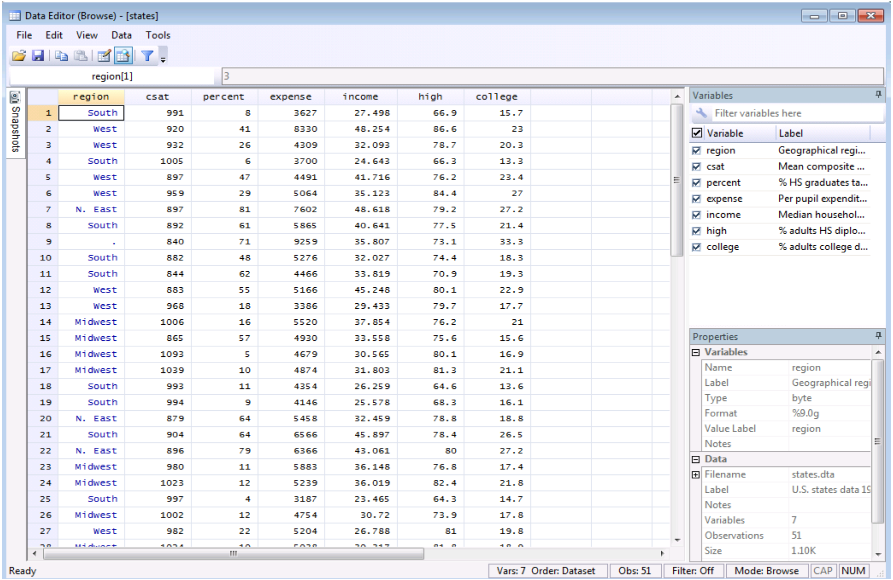
$$\frac{\hat{\boldsymbol{\beta}}_2' x_2' x_2 \hat{\boldsymbol{\beta}}_2 / (k-1)}{e'e/(n-k)} = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k).$$

Empirical Example

Example

Are SAT scores higher in states that spend more money on education controlling for other factors?

- Outcome variable (Y): SAT scores in each state (`csat`)
- Explanatory varibles (X):
  - per pupil expenditures primary & secondary (`expense`)
  - % HS graduates taking SAT (`percent`)
  - median household income (`income`)
  - % adults with HS diploma (`high`)
  - % adults with college degree (`college`)
  - region (`region`)



	region	csat	percent	expense	income	high	college
1	South	991	8	3627	27.498	66.9	15.7
2	West	920	41	8330	48.254	86.6	23
3	West	932	26	4309	32.093	78.7	20.3
4	South	1005	6	3700	24.643	66.3	13.3
5	West	897	47	4491	41.716	76.2	23.4
6	West	959	29	5064	35.123	84.4	27
7	N. East	897	81	7602	48.618	79.2	27.2
8	South	892	61	5865	40.641	77.5	21.4
9	.	840	71	9259	35.807	73.1	33.3
10	South	882	48	5276	32.027	74.4	18.3
11	South	844	62	4466	33.819	70.9	19.3
12	West	883	55	5166	45.248	80.1	22.9
13	West	968	18	3386	29.433	79.7	17.7
14	Midwest	1006	16	5520	37.854	76.2	21
15	Midwest	865	57	4930	33.558	75.6	15.6
16	Midwest	1093	5	4679	30.565	80.1	16.9
17	Midwest	1039	10	4874	31.803	81.3	21.1
18	South	993	11	4354	26.259	64.6	13.6
19	South	994	9	4146	25.578	68.3	16.1
20	N. East	879	64	5458	32.459	78.8	18.8
21	South	904	64	6566	45.897	78.4	26.5
22	N. East	896	79	6366	43.061	80	27.2
23	Midwest	980	11	5883	36.148	76.8	17.4
24	Midwest	1023	12	5239	36.019	82.4	21.8
25	South	997	4	3187	23.465	64.3	14.7
26	Midwest	1002	12	4754	30.72	73.9	17.8
27	West	982	22	5204	26.788	81	19.8

SAT Scores Example

```
. regress csat expense, robust

Linear regression                               Number of obs =      51
                                                F(  1,    49) =   36.80
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2174
                                                Root MSE      =  59.814
```

csat	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
expense	-.0222756	.0036719	-6.07	0.000	-.0296547	-.0148966
_cons	1060.732	24.35468	43.55	0.000	1011.79	1109.675

SAT Scores Example

```
Linear regression                               Number of obs =        51
                                                F(  5,     45) =     50.90
                                                Prob > F      =    0.0000
                                                R-squared     =    0.8243
                                                Root MSE      =    29.571
```

csat	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
expense	.0033528	.004781	0.70	0.487	-.0062766	.0129823
percent	-2.618177	.2288594	-11.44	0.000	-3.079123	-2.15723
income	.1055853	1.207246	0.09	0.931	-2.325933	2.537104
high	1.630841	.943318	1.73	0.091	-.2690989	3.530781
college	2.030894	2.113792	0.96	0.342	-2.226502	6.28829
_cons	851.5649	57.28743	14.86	0.000	736.1821	966.9477

SAT Scores Example

- Suppose we would like to test wether two coefficients are jointly different from zero
- We can use `test` command in STATA
- Specifically, we test the null hypothesis that both coefficients (high and college) do not have any effect on SAT scores
- `test high college`
  (1) high $= 0$
  (2) college $= 0$
  F $(2, 40) = 17.12$
  Prob $>$ F $= 0.0000$
- The *p*-value is 0.0000, we reject the null and conclude that both variables have indeed a significant effect on SAT scores

Restricted Least Squares

- Sometimes we have prior information from economic theory, which may impose a certain restriction on the regression coefficients.
- One may do estimation and testing subject to the prior information.

Example

Consider a simple model of investment.

$$\ln I_t = \beta_1 + \beta_2 i_t + \beta_3 \Delta p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t,$$

where i_t is the nominal interest rate, Δp_t is the rate of inflation and $\ln Y_t$ is the log output. An alternative theory says "Investors care only about real interest rates". The testable model becomes

$$\ln I_t = \beta_1 + \beta_2 (i_t - \Delta p_t) + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t,$$

which implies the restriction $\beta_2 + \beta_3 = 0$.

Restricted Least Squares

- To proceed the restricted least square, we adopt to the Lagrange method.
- The objective function is:

$$\min_{\beta} (Y - X\beta)' (Y - X\beta) \quad \text{subject to } R\beta = q.$$

- Lagrangian is:

$$\mathcal{L}(\beta, \lambda) = (Y - X\beta)' (Y - X\beta) - 2\lambda' (R\beta - q).$$

- FOCs are straightforward.

Restricted Least Squares

- FOCs:

$$\frac{\partial \mathcal{L}(\beta,\lambda)}{\partial \beta} = -2X'Y + 2\left(X'X\right)b^* - 2R'\lambda^* = 0 \tag{2}$$

$$\frac{\partial \mathcal{L}(\beta,\lambda)}{\partial \lambda} = Rb^* - q = 0 \tag{3}$$

- Premultiplying (2) by $R\left(X'X\right)^{-1}$ provides the solution of λ^*.

$$\lambda^* = (R\left(X'X\right)^{-1}R')^{-1}(q - R\hat{\beta}). \tag{4}$$

- Plugging (4) into (2) gives the solution for restricted least square estimator b^*. (in terms of the unrestricted least square estimator $\hat{\beta}$)

$$b^* = \hat{\beta} + \left(X'X\right)^{-1}R'(R\left(X'X\right)^{-1}R')^{-1}(q - R\hat{\beta}).$$

Restricted Least Squares – Unbiasedness

Theorem

$E[b^*] = \beta$

Proof.

$$\mathrm{E}[b^*] = \mathrm{E}[\hat{\beta} + (X'X)^{-1} R'(R\,(X'X)^{-1} R')^{-1}(q - R\hat{\beta})] = \beta.$$

$\square$

Restricted Least Squares – Variance

Theorem

$$var[b^*] = \sigma^2 \left(X'X\right)^{-1} - \sigma^2 \left(X'X\right)^{-1} R'(R\left(X'X\right)^{-1} R')^{-1} R\left(X'X\right)^{-1}$$

Proof.

Write $b^* = (I - \left(X'X\right)^{-1} R'(R\left(X'X\right)^{-1} R')^{-1} R)\hat{\beta} + Cq = T_1\hat{\beta} + T_2 q$

$$\begin{aligned} \text{var}[b^*] &= T_1 \text{var}[\hat{\beta}] T_1' \\ &= T_1 \sigma^2 \left(X'X\right)^{-1} T_1' \\ &= \sigma^2 \left(X'X\right)^{-1} - \sigma^2 \left(X'X\right)^{-1} R'(R\left(X'X\right)^{-1} R')^{-1} R\left(X'X\right)^{-1}. \end{aligned}$$

□

Restricted Least Squares – Efficiency

Theorem

Under the restriction,

$$b^*|X \sim \mathcal{N}(\beta, \sigma^2 \left(X'X\right)^{-1} - \sigma^2 \left(X'X\right)^{-1} R'(R \left(X'X\right)^{-1} R')^{-1} R \left(X'X\right)^{-1}).$$

Proof.

Combined the results from above two theorems. □

Restricted Least Squares – Efficiency

Theorem

b^* is BLUE, i.e., $var[b^*] \leq var[\hat{\beta}]$.

Proof.

$\text{var}[\hat{\beta}] - \text{var}[b^*] = \sigma^2 (X'X)^{-1} R'(R(X'X)^{-1} R')^{-1} R(X'X)^{-1}$, which is a p.s.d. matrix. [intuition?] □

Restricted Least Squares

- There is an alternative view to see the test for restrictions.
- Idea is to compare the unrestricted RSS and restricted RSS.
- Note that the restricted OLS residual is

$$
\begin{aligned}
e^* = Y - Xb^* = Y - X\hat{\beta} + X\hat{\beta} - Xb^* &= e + X(\hat{\beta} - b^*) \\
&= e + X\left(X'X\right)^{-1} R'(R\left(X'X\right)^{-1} R')^{-1}(R\hat{\beta} - q).
\end{aligned}
$$

$$
\begin{aligned}
e^{*\prime}e^* &= e'e + (\hat{\beta} - b^*)'X'X(\hat{\beta} - b^*) \geq 0(?) \\
&= e'e + (R\hat{\beta} - q)'(R\left(X'X\right)^{-1} R')^{-1}(R\hat{\beta} - q) \\
e^{*\prime}e^* - e'e &= (R\hat{\beta} - q)'(R\left(X'X\right)^{-1} R')^{-1}(R\hat{\beta} - q),
\end{aligned}
$$

which is part of the numerator in F statistics.

Restricted Least Squares

- Hence, the F statistics could be rewritten as

$$F = \frac{\left(e^{*\prime}e^{*} - e^{\prime}e\right)/J}{e^{\prime}e/\left(n-k\right)} \tag{5}$$

$$= \frac{\left(RSS_r - RSS_u\right)/J}{RSS_u/\left(n-k\right)} \tag{6}$$

$$= \frac{\left(R_u^2 - R_r^2\right)/J}{\left(1-R_u^2\right)/\left(n-k\right)}. \tag{7}$$

Restricted Least Squares

Example

Going back to Example of the overall test, we could find $R_r^2 = 0$ since $RSS_r = TSS$. [check it!] Using this fact and plugging into (6), we have the overall test of regression being

$$F = \frac{R_u^2 / (k-1)}{(1 - R_u^2) / (n-k)} \sim F(k-1, n-k).$$

In this case, we don't need to estimate the restricted model.